



圖一 國立故宮博物院每日絡繹不絕的遊客 作者攝

# 博物館觀眾研究新契機

## 大數據

賴鼎陞

近年來「大數據」已成為各媒體上熱門的話題，本文探討相關分析方法應用在博物館觀眾研究的發展，並以一個案例研究，進行國立故宮博物院參觀人數的構成分析，以及未來訪客人數的預測。研究發現，有別於問卷調查、訪談等傳統觀眾研究方法，經常用以發現觀眾對過去展覽、服務等方面的意見；而大數據分析卻能揭露一些內隱的資訊，以及預測未來趨勢，使研究面向能更多元化。

### 大數據議題的興起

近年來傳播媒體經常報導「大數據(Big Data)」的相關議題，一般而言，是指涉及的資料數量龐大，而且複雜，無法以現有系統直接處理，包括：擷取(capture)、策畫(curation)、儲存(storage)、搜尋(search)、分享(sharing)、轉換(transfer)、分析(analysis)和視覺化(visualization) (註)

一、特別是在商業行銷、工業生產、資訊傳播、公共行政、醫療保險等不同領域，都有實務案例。

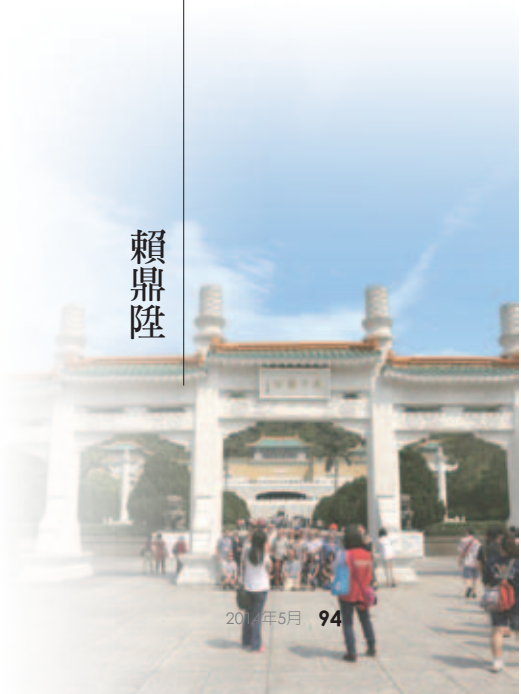
「大數據」一詞大約出現在一九九八年間，因國際網路全球風行，各種數位資料化不斷產生，並且透過網路大量流通，資、通訊產業針對大量不易處理的資料，所給予的一個通用稱呼。然而，經過十餘年的產

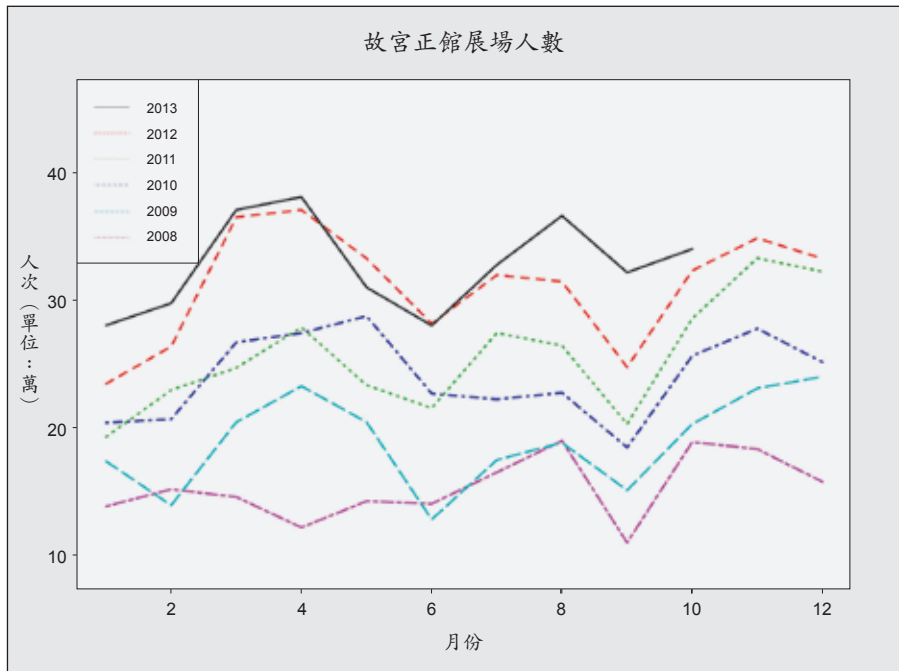
業技術發展，資料存取方式從本地到雲端、從桌上到行動、從公用到個人等不斷地躍進，為了有效地分析不斷產出的巨量資料，進而推衍出「商業智慧」，「大數據分析(Big Data Analytics)」也順勢成為專業人員學習的新興技術能力。(註二)

在眾多新興議題中，筆者認為大數據領域專家Mayer-Schonberger與

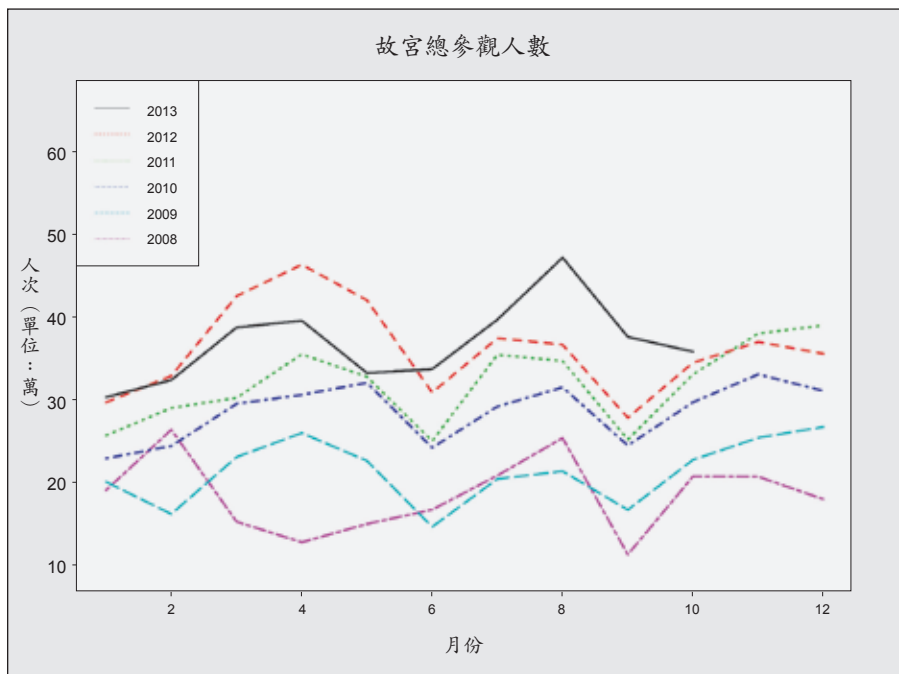
Cukier提出的批判性觀點，最具原創性，而且適用於不同的專業領域。他們認為：「過去許多無法測量、儲存、分析的事務，現在都已經開始資料化，在巨量資料時代，我們需要新的準則，其將挑戰我們的生活、工作和思維，以及與世界互動的方式。」

首先，他們認為在過去資料不足的時代，由於技術限制，很難掌握到全體的資料，只好抽取小量樣本，並使用各種統計技巧去推估母體的概況。但是在巨量資料時代，掌握全體的資料已不是夢，不用再畫地自限，拘泥於統計觀點，要大膽而廣泛的蒐羅所有的資料，這是各行各業都適用的原則。其次，巨量資料的來源與格式不一，但因為資料數量遠比資料品質更重要，即便資料較雜亂，也能過濾、萃取出資料的精華，顯露真正意涵。再者，過去的統計方法注重因果關係的推論，而大數據分析，則以於發現「相關性」為重點，所以不必強求探究「因果關係」；換言之，研究者不必追根究底地追查「為何如此(know why)」，只要知道「正是如此





圖二 過去六年（2008/1~2013/10）故宮正館展場人數



圖三 過去六年（2008/1~2013/10）故宮總人數

展場服務等項目的意見。另外，也可能在某一特展的籌備期間，邀請焦點團體進行展前評量座談，瞭解觀眾對展覽的期待；以及在辦理各項教育活動之後，發放問卷以

評估活動績效。這些都是在某個時間區間內進行資料蒐集，以獲得資料樣本，進而推論母體的概況。而上述的抽樣調查方式，是局限在一段特定時間內蒐集資料，藉由小

量的樣本進行統計分析，以推估母體的概況，這種研究方法被數學與統計學家稱為「橫斷面研究（Cross-sectional Research）」，例如：辦理某項特展後，調查是否提昇整體滿意度；或是

以目前資料顯示，相較於在工商界已經發展的一些成功案例，大數據分析在博物館領域，各項議題仍屬在初步問題定位階段，尚未實際施行。部分的小型實驗計畫，也還

績效等。（註六）

舉例來說，依據相關研究顯示，「Y世代」的年輕族群中（意指大約在一九八二到二〇〇〇年間出生者），在三個人當中，就有二個人覺得「博物館對於他們的休閒活動，有很好的價值」，由於網路媒體在他們的成長歷程扮演重要的角色，他們喜好透過這樣的管道溝通、傳遞訊息，對於「口碑經驗」的認同感，比其他

### 從橫斷面研究，擴展到縱斷面研究

一般而言，博物館為進行觀眾研究，通常會採用問卷調查、訪談等方式進行調查，舉例來說，國立故宮博物院於每年都辦理觀眾滿意度調查工作，依據訪客結構（如：男性或女性、團體或散客、家庭或年輕族群等）進行抽樣，調查進行方式，則是透過擇期於展場進行問卷調查、或者電話訪談，以瞭解觀眾對展覽內容、

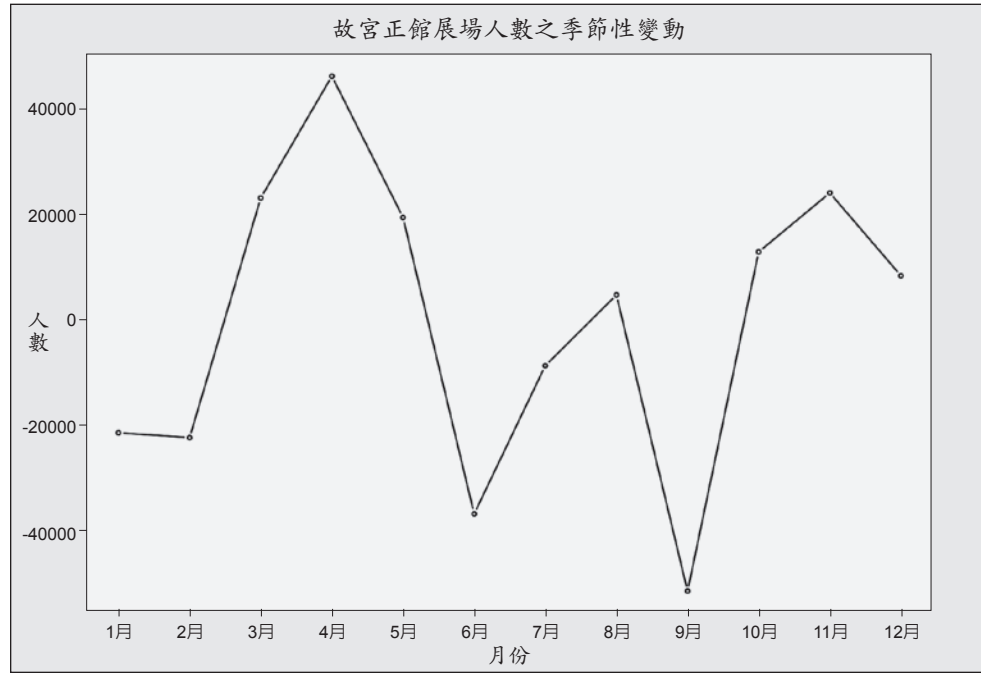
（know what）」，就已經足夠研究與應用所需了。（註三）

### 大數據分析在博物館領域的發展

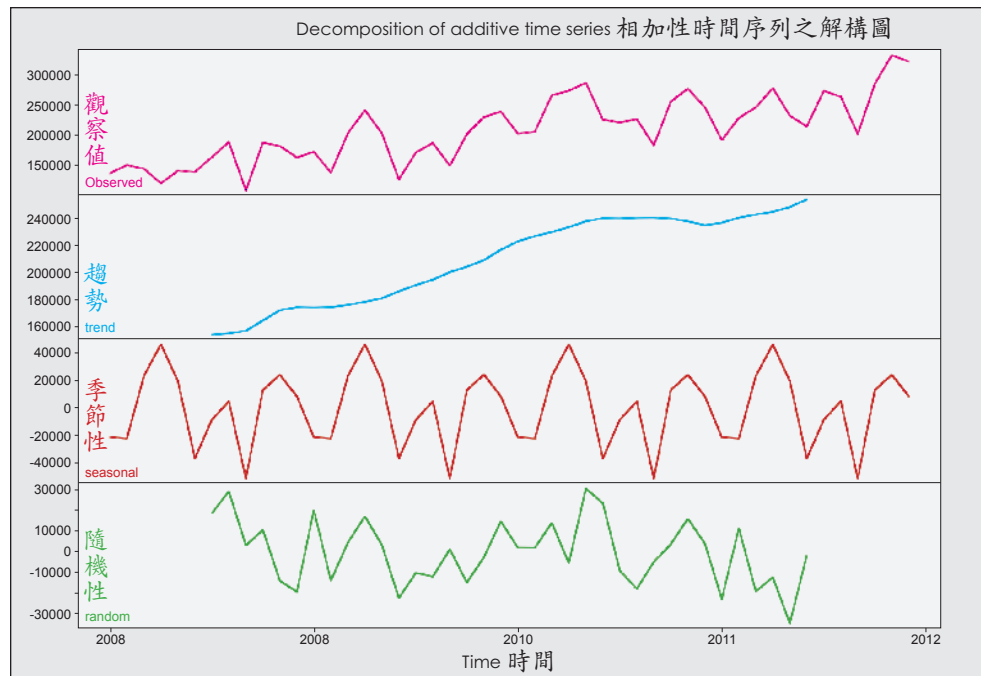
最近幾年，在國際博物館領域，大數據分析（Big Data Analytics）也逐漸受到關注，在各專業討論社群中，越來越多專業人士以專文探討大數據分析在博物館的發展。（註四）

未發佈具體成果。這可能顯露出一個共通的現象，許多博物館、美術館等文化機構，雖然具備進行大數據研究的潛力，但目前仍受到一些限制，包括：缺乏歷史資料、沒有研究人員和經費，以及原本設定的營運目標，並未涵蓋進行量化資料的研究與分析。（註七）

因此，筆者認為，博物館除了專注於傳統價值的產出之外，例如：展覽、導覽等，也可透過社交媒體進行傳播，發展具故事性的行銷活動，以得到更多不同族群的認同。而發現這些決策的管道，除了仰賴博物館人員的經驗之外，針對訪客或網站歷史資料進行大數據分析，是最有系統性的方法之一。



圖五 故宮總人數之季節性變動趨勢



圖六 故宮總人數之時間序列圖形

在資料探勘的工具方面，我們選擇以「R 語言」進行程式設計，其源自於紐西蘭奧克蘭大學統計學系的

計畫，目前是由軟體核心團隊來維護，提供許多關於統計、繪圖等軟體套件。

### 主要發現

在二〇一三年一〇月，軟體程式開發完成後，專案則著手進行資料



圖四 每年春天是故宮人潮高峯期。 作者攝於2014年4月17日

### 案例研究：以故宮大數據分析為例

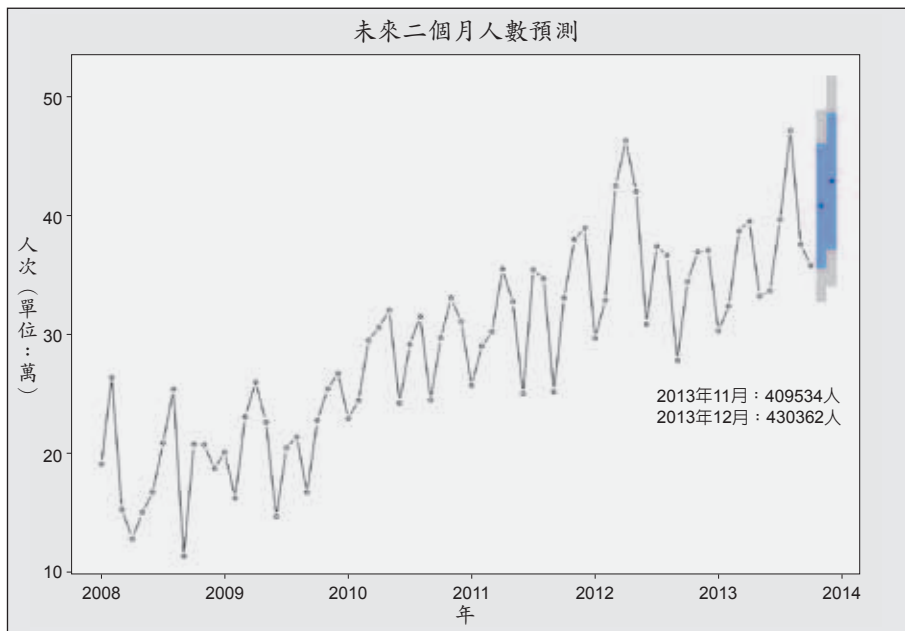
提供種服務項目之後，調查是否提高到訪意願等。此方法花費時間、經費較少，但缺點是博物館無法瞭解、整體發展過程的變化。

若是採用「橫斷持續研究 (Cross-sectional Research)」方式，延伸抽樣期間、增加抽樣次數，例如：由每年一次改為每年四次，雖然可以適當補足上述弱點，但人力成本勢必倍增；但無論如何，抽樣資料依然是非連續性的，仍然無法探究博物館服務的發展過程。

而本文所討論的「大數據分析」，即屬於「縱斷面研究 (Longitudinal Research)」方法，以連續性的抽樣區間，持續收集資料，除了可以補足上述縱斷面研究的不足，讓博物館服務的發展趨勢得以揭露；而大量數據中，可能潛藏一些讓博物館管理者和訪客都不易發覺的訊息，更可運用「資料探勘 (Data Mining)」的技術進行探究。

在博物館界，各項與觀眾相關的研究中，大數人最關切的議題，莫過於觀眾人數、觀眾組成、觀眾行為等調查。依據英國《藝術報》在二〇一三年四月所公布的資料顯示，在全世界藝術類博物館中，國立故宮博物院的總參觀人數，連續兩年位居第七（三八四・九、四三六萬），是在排行榜前十名之內，唯一的亞洲博物館。（圖一）而事實上，自二〇〇八年七月開放大陸居民來臺觀光之後，故宮訪客人數逐年成長，究竟故宮人數成長的趨勢為何？這些數據背後，是否隱藏我們尚未得知的訊息？這些問題，引起了我們的研究興趣。

因此，一個「大數據分析」的專案研究在故宮因運而生，目標是利用資料探勘的相關技術，結合包括資料視覺化 (data visualization)、統計 (statistics)、資料庫 (database) 等多種資料處理方式，針對故宮二〇〇八年以後的參觀人數資料進行分析，以期從大數據中，獲得更多背後的隱涵的故事，例如：樣式 (pattern)、關聯 (association) 等。



圖八 以歷史資料(2008/1~2013/10)預測未來二個月(2013/11~12)人數

光景點等也可適用，特別在展示、教育、推廣等觀眾行為模式上，文教、觀光等領域有許多共通之處，因此，筆者可以總結出幾項研究發現：

一、「資料視覺化」是在大數據分析

二、以觀光旅遊為主要目標的觀眾群，例如：旅行社安排的團體，參觀人數資料會呈現季節性時間序列的特徵，此一現象，也合理地歸因於觀光產業的供需平衡問題，包括：交通、住宿、領隊和導遊數等。因此，博物館進行大數據分析，應該結合外部的相關數據，例如：交通部觀光局、台北市觀光傳播局、或者同一旅程中其他知名景點等資料，以探究博物館在觀光功能中，所扮演的角色。

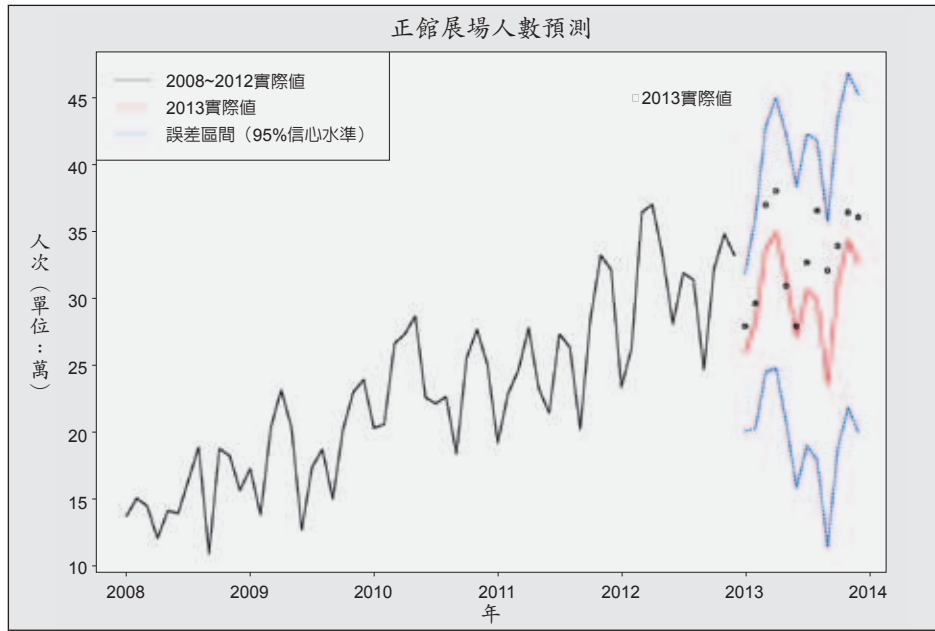
三、大數據分析所採行的縱斷面研究方法，和博物館經常使用的問卷

很重要的研究方法之一，在一般年度報告中常見的「成長率」計算，例如：「今年的某一月份，相較去年同期，成長率為百分之幾」，並不容易呈現長期發展走向，以及各細節資料之間的相關性。資料視覺化雖無法提供精準的數據，但可提供管理者更容易閱讀的資訊，以進行適當的決策。

四、博物館大數據資料的收集，除了到訪觀眾的背景資料，亦可納入博物館媒體的使用情況，包括：網站點閱、檢索系統存取紀錄，臉書社交媒體互動等觀眾行為的收集。

五、博物館辦理各項特展、以及教育性活動，明顯有助於增加博物館參觀的人數，提高非團體訪客參觀意願、或觀眾再訪比率。

以故宮為例，研究顯示每年三至五月，國外訪客將達到季節性高峰，而國內訪客較多集中在寒暑假、大型



圖七 以過去五年(2008~2012)歷史資料預測2013年人數

而本文完稿時(二〇一三年三月)的資料顯示，二〇一三年故宮總人數

透過上述案例研究可以推論，在故宮進行「大數據分析」的經驗，在其他博物館、美術館、動物園、觀

接著我們透過ARIMA模型進行未來預測，依據過去五年(二〇〇八年至二〇一二年)歷史資料計算，推估下一年(二〇一三年)的人數走勢，而在百分之九十五的信心水準內，推算出一條紅色線，代表未來預測走勢，以及兩條藍色的上、下誤差界限，而黑色線是依據歷史資料繪製實際走勢。(圖七)

結語 透過上述案例研究可以推論，在故宮進行「大數據分析」的經驗，在其他博物館、美術館、動物園、觀

探勘工作。透過資料視覺化的呈現，顯見過去六年(二〇〇八年一月至二〇一三年十月)故宮正館展場人數逐年增加，年度參觀人數曲線層層而上

利用「時間序列分析(Time Series Analysis)」的相關研究方法(註八)，我們可以得知，近年來，故宮每月參觀人數曲線，呈現季節性的變化趨勢。每年有三個主要的高峰期，最高峰是三至五月(圖四)、次高峰是十至十二月，第三高是七至八月。(圖五)

除此之外，我們當時以過去歷史資料(二〇〇八年一月至二〇一三年十月)推估未來二個月的故宮人數，在百分之九十五的信心水準內，我們預測二〇一三年十一月是四十九萬九千五百三十四人次，上下區間為三十二萬八千七百六十一到四十九萬三千零五人；而二〇一三年十二月是四十三萬三千六百二十人次，上下區間為三十四萬一千七百五十八到五十一萬八千九百六十五。而實際上，經過事後查證：二〇一三年十一月是四十一萬六千六百七十七人，二〇一三年十二月是三十九萬七千二百八十八人，顯示皆在預測模型的可信範圍以內。(圖八)

# 履踪

In Their Footsteps

A Special Exhibition of Images and Documents on Indigenous Peoples in Taiwan

2013 12/14 - 2014 5/19

臺灣原住民  
文獻圖畫特展

《履踪：臺灣原住民文獻圖特展》圖錄於本院禮品部販售



11143 台北市士林區至善路二段221號 No.221, Sec. 2, Zhishan Rd., Shilin Dist., Taipei City 11143, Taiwan (R.O.C.) Tel:02-6610-3600 Fax:02-2882-1440 http://www.npm.gov.tw

第一展廳圖錄列室 Gallery 104



展覽、文獻館特展等特定期間，因此，故宮可以在展示、導覽、服務、動線、活動、文創商品等各方面預先做規劃，讓不同背景的觀眾，可以明顯地感受到特別的「博物館經驗」，而不僅受限於對故宮的刻板印象。

從管理層面來看，也較能在有限的人力、場域、經費等限制之下，對於故宮未來的服務創新，提出更積極、有效的方案。而大數據分析技術，適可在決策過程中扮演「資訊幕僚」的角色。

展望未來，適當引進現代化的技術，可大幅減低人工處理的負擔，並提昇「大數據」的完整性，例如：票券電子化、人流與車流自動計數、觀眾意見調查網路化等，都是現今已成熟、容易導入應用的工具。

綜上所述，未來臺灣的文化、教育、觀光領域，需要許多單位的大數據的互相連結，以及整合分析，才能逐步建構更完整的觀眾行為模型，例如：博物館、藝文活動、觀光旅遊等資料的彙集，對於非營利教育傳播、

文創產業的行銷，才能產生群聚的效益。雖然相關研究才剛起步，但可以確定的是，在大數據分析的思維和研究方法上，與上述Mayer-Schonberger與Cukier提出觀點是相通的，本文實證了「大數據分析」在博物館觀眾研究的初步價值，也開啓了一個新的方向。

筆者感謝中華R軟體學會鄧慶士博士、李明昌博士、曾建祥先生、林明輝先生在相關演算法和R語言程式設計上，給予本專案的協助。

作者任職於本院教育展覽處

註釋

1. "Big Data". At [http://en.wikipedia.org/wiki/Big\\_data](http://en.wikipedia.org/wiki/Big_data)
2. Gill Press (2013). "A Very Short History Of Big Data". Forbes". <http://www.forbes.com/sites/gillpress/2013/05/09/a-very-short-history-of-big-data/>
3. 麥爾哥伯格、庫基耶著，林俊宏譯，《大數據》，臺北：天下文化，二〇一三年五月。譯者Viktor Mayer-Schonberger & Kenneth Cukier. Big Data: A Revolution That Will Transform How We Live, Work, and Think. Eamon Dolan/Houghton Mifflin Harcourt. 2013.05.
4. (a) Suse Cairns (2012). "One Provocation for Big Data" - <http://museumgeek.wordpress.com/2012/02/01/one-provocation-for-big-data/>; (b) Xinxin Guo (2013). "Big Data and Museums: Wave of the Future?" - <http://nyuisva.wordpress.com/2013/04/15/big-data-and-museums-wave-of-the-future/>; (c) John Lucas (2013). "From Zoos to Museums, Cultural Venues Can Learn a Lot from Big Data" - <http://asmarterplanet.com/blog/2013/06/from-zoos-to-museums-cultural-venues-can-learn-a-lot-from-big-data.html>
5. "Big Data/Small Data: GLAM Collections in the 21st Century" - <http://mw2013.museumandtheweb.com/proposals/big-data-small-data-glam-collections-in-the-21st-century/>

6. "Big Data: Frontier" - <http://www.aam-us.org/events/annual-meeting/program/sessions-and-events?id=2333>
7. Rachel schutt (2013). "The Data Science of Art" - <http://columbiadatascience.com/2012/10/03/the-data-science-of-art/#comments>
8. 林茂文，《時間數列分析與預測：管理與財經之應用》，臺北：華泰文化，二〇〇六年六月。

9. "Big Data: Frontier" - <http://www.aam-us.org/events/annual-meeting/program/sessions-and-events?id=2333>